Private Queries in Location-Based Services: Anonymizers are Not Necessary

Gabriel Ghinita\textsuperscript{1} Panos Kalnis\textsuperscript{1}
Ali Khoshgozaran\textsuperscript{2} Cyrus Shahabi\textsuperscript{2}
Kian Lee Tan\textsuperscript{1}

\textsuperscript{1} National University of Singapore
\textsuperscript{2} University of Southern California

Location-Based Services (LBS)

- LBS users
  - Mobile devices with GPS capabilities
- Queries
  - NN Queries
  - Location server is NOT trusted

Problem Statement

- Queries may disclose sensitive information
  - Query through anonymous web surfing service
- But user location may disclose identity
  - Triangulation of device signal
  - Publicly available databases
  - Physical surveillance
- How to preserve query source anonymity?
  - Even when exact user locations are known

PIR Overview

- Computationally hard to find $i$ from $q(i)$
- Bob can easily find $X_i$ from $r$ (trap-door)

Existing LBS Privacy Solutions

Spatial K-Anonymity

- Query issuer "hides" among other K-1 users
  - Probability of identifying query source $\leq 1/K$
  - Idea: anonymizing spatial regions (ASR)
Casper\cite{Mok06}

- Quad-tree based
  - Fails to preserve anonymity for outliers
  - Unnecessarily large ASR size

\begin{itemize}
  \item Let \( K = 3 \)
  \item If any of \( u_1, u_2, u_3 \) queries, ASR is \( A_1 \)
  \item If \( u_4 \) queries, ASR is \( A_2 \)
  \item \( u_4 \)'s identity is disclosed
\end{itemize}

\cite{Mok06} - Mokbel et al, The New Casper: Query Processing for Location Services without Compromising Privacy, VLDB 2006

Reciprocity

\begin{itemize}
  \item A
  \item B
  \item C
  \item D
  \item E
  \item F
\end{itemize}

\cite{CM07} - C.-Y. Chow and M. Mokbel "Enabling Private Continuous Queries For Revealed User Locations". In Proc. of SSTD 2007

Hilbert Cloak (HC)

- Based on Hilbert space-filling curve
  - Index users by Hilbert value of location
  - Partition Hilbert sequence into "K-buckets"

\begin{itemize}
  \item Let \( K = 3 \)
  \item If any of \( u_1, u_2, u_3 \) queries, ASR is \( A_2 \)
  \item If \( u_4 \) queries, ASR is \( A_1 \)
  \item \( u_4 \)'s identity is disclosed
\end{itemize}

\cite{Mok06} - Mokbel et al, The New Casper: Query Processing for Location Services without Compromising Privacy, VLDB 2006

Continuous Queries\cite{CM07}

- Problems
  - ASRs grows large
  - Query dropped if some user in \( U \) disconnects

\cite{CM07} - C.-Y. Chow and M. Mokbel "Enabling Private Continuous Queries For Revealed User Locations". In Proc. of SSTD 2007

Space Encryption\cite{KS07}

- Drawbacks
  - Answers are approximate
  - Makes use of tamper-resistant devices
  - May be vulnerable if some POI are known


Motivation

- Limitations of existing solutions
  - Assumption of trusted entities
    - Anonymizer and trusted, non-colluding users
  - Considerable overhead for sporadic benefits
    - Maintenance of user locations
  - No privacy guarantees
    - Especially for continuous queries
Our Approach

PIR Theoretical Foundations

- Let $N = q_1 * q_2$, $q_1$ and $q_2$ large primes
  
  $\mathbb{Z}_N^* = \{ x = x \mod (N, r) = 1 \}$
  
  $QR = \{ y \in \mathbb{Z}_N^* | \exists x \in \mathbb{Z}_N^* : y = x^2 \mod N \}$

- Quadratic Residuosity Assumption (QRA)
  - QR/QNR decision computationally hard
  - Essential properties:
    
    $QR * QR = QR$
    $QR * QNR = QNR$

PIR Protocol for Binary Data

$$z_j = \prod_{i=1}^{4} \frac{y_{i,j}}{x_{i,j}}$$

Approximate Nearest Neighbor

- Data organized as a square matrix
  - Each column corresponds to index leaf
  - An entire leaf is retrieved – the closest to the user

Exact Nearest Neighbor

- Only $z_2$ needed

LBS Privacy with PIR

- Two-party cryptographic protocol
  - No trusted anonymizer required
  - No trusted users required

- No pooling of a large user population required
  - No need for location updates

- Location data completely obscured
Avoiding Redundant Computations

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>__</td>
<td>__</td>
<td>__</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- Data mining
  - Identify frequent partial products

Parallelize Computation

- Values of $z$ can be computed in parallel
- Offline phase: master scatters PIR matrix
- Online phase:
  - Master broadcasts $y$
  - Each worker computes $z$ values for its strip
  - Master collects $z$ results

Experimental Settings

- Sequoia dataset + synthetic sets
  - 10,000 to 100,000 POI
- Modulus up to 1280 bits

Computation/Communication Overhead (Approximate)

Parallel Execution
Conclusions

- PIR-based LBS privacy
  - No need to trust third-party
  - Secure against any location-based attack
- Future work
  - Further reduce PIR overhead
  - Support more complex queries
  - Include more POI information in the reply

Discussion

- Given the parallelization, compression, multiplication reduction, rectangular shape $M_i$, how much is communication/computation saved?
- How do you compare the previous two approaches?
- What do you think is the major challenge in achieving privacy-aware LBS?

Reciprocity

- Consider querying user $u_q$ and ASR $A_q$
- Let $AS_q = \{\text{set of users enclosed by } A_q\}$
- $A_q$ has the reciprocity property iff
  1. $|AS| \geq K$
  2. $\forall u_i, u_j \in AS, u_i \in AS_j \land u_j \in AS_i$

Continuous Queries\cite{CM07}

- Extends reciprocity to moving clients
- Let $A_{t_0}$ be ASR at time $t_0$, let $U$ be the users in $A_{t_0}$
- At time $t$, ASR is MBR of $U$ (at new locations)
- Problems
  - ASR grows large
  - Query dropped if some user in $U$ disconnects

---


\cite{CM07} C.-Y. Chow and M. Mokbel “Enabling Private Continuous Queries For Revealed User Locations”. In Proc. of SIGMOD 2007
Space Encryption [KS07]

- Does not employ SKA
  - each POI is mapped to 1-D value (Hilbert)
  - fractal parameters are kept secret
  - answers are approximate
  - makes use of tamper-resistant devices
  - may be vulnerable if some POI are known


Rectangular PIR Matrix

Server Computation Overhead

Approximation Error

Bibliography


http://anonym.comp.nus.edu.sg